Proposal: Music Genre Classification

Mark Gameng

Illinois Institute of Technology

mgameng1@hawk.iit.edu

1. Description

Music, throughout the years, have been skyrocketing in terms of music generation and have also resulted in new styles/genres of music. Music platforms in general are producing thousands of new songs each day, with Spotify having listed over 50 million songs and over 40 thousand new songs are added every day. Years ago, the genre of songs were classified manually by people with musical knowledge. With the amount of music generation in our current era, it is imperative that music genre classification be automated through musical analysis and machine learning. Automated music genre classification will allow platforms like Spotify, Pandora, Soundcloud to better serve its customers by having a more accurate search and recommender systems.

1.1. Proposal

Due to the amount of growth in the music industry throughout the years and the amount of data that can be extracted from an audio, interest in the field of Music Information Retrieval (MIR) has been increasing. Specifically, one popular topic that has garnered countless studies in MIR is music genre classification. As a result, there has also been many datasets containing songs with information like genre. With machine learning, computer scientists have been trying to tackle the automation of music genre classification.

Few of the datasets that are majorly used in these type of studies are GTZAN [6] and Extended Ballroom [1]. GTZAN is the most used public dataset which consists of 1000 audio clips that are 30 seconds each and are equally classified to 10 different genres. The Extended Ballroom dataset consists of 4180 audio clips that are 30 seconds each which are from 13 different ballroom genres. Other datasets to note are Free Music Archive (FMA) and Million Song which are much larger datasets with 0.1 and 1 million clips respectively.

Currently, the best result that I know of got 92% and 92.5% accuracy on GTZAN and extended ballroom dataset respectively. That was done using a parallel recurrent convolutional neural network (PRCNN) [7]. Other notable methods are CNN [9], KCNN + SVM [8], and DNN [4], which got 87.4%, 83.9%, and 83% respectively. Other works also used lyrics [5], album picture either separately or in combination with audio to determine the genre [3].

I propose trying to experiment different things such as DNN, CNN, RNN, SVM, etc., on different feature extractions and seeing what works and which is best. Recently, there might also be new feature extractions and techniques that may be useful such as graph-based feature extraction [2] and more. So, I will be trying different combinations of feature extractions as well as parallel/combinations of models to determine the genre.

1.2. Preliminary Plan

1. Decide which dataset to use. Either both GTZAN and extended ballroom dataset because they are widely used. Can also use FMA or million song dataset which have much larger sample size

2. Extract features from audio clips such as MFCC, mel-spectrum, SFM, STFT, etc.

3. Implement models and experiment with it

4. Look at the results and modify or optimize the models. My goal would be to get an accuracy of at least 70%

References

[1] Ugo Marchand and Geoffroy Peeters. The extended ballroom dataset. 2016.

[2] Dirceu de Freitas Piedade Melo, Inacio de Sousa Fadigas, and Hernane Borges de Barros Pereira. Graph-based feature extraction: A new proposal to study the classification of music signals outside the time-frequency domain. *Plos one*, 15(11):e0240915, 2020.

[3] Sergio Oramas, Oriol Nieto, Francesco Barbieri, and Xavier Serra. Multi-label music genre classification from audio, text, and images using deep features. *arXiv preprint arXiv:1707.04916*, 2017.

[4] Siddharth Sigtia and Simon Dixon. Improved music feature learning with deep neural networks. In *2014 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 6959–6963. IEEE, 2014.

[5] Alexandros Tsaptsinos. Lyrics-based music genre classification using a hierarchical attention network. *arXiv preprint arXiv:1707.04678*, 2017.

[6] George Tzanetakis and Perry Cook. Musical genre classification of audio signals. *IEEE Transactions on speech and audio processing*, 10(5):293–302, 2002.

[7] Rui Yang, Lin Feng, Huibing Wang, Jianing Yao, and Sen Luo. Parallel recurrent convolutional neural networks-based music genre classification method for mobile devices. *IEEE Access*, 8:19629–19637, 2020.

[8] Pengjing Zhang, Xiaoqing Zheng, Wenqiang Zhang, Siyan Li, Sheng Qian, Wenqi He, Shangtong Zhang, and Ziyuan Wang. A deep neural network for modeling music. In *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*, pages 379–386, 2015.

[9] Weibin Zhang, Wenkang Lei, Xiangmin Xu, and Xiaofeng Xing. Improved music genre classification with convolutional neural networks. In *Interspeech*, pages 3304–3308, 2016.